

修士論文の和文要旨

大学院情報システム学 研究科		博士前期課程	情報システム設計学	専攻
氏名	林 和宏		学籍番号	0550030
論文題目	多次元データマイニングによる Web アーカイブの 構造分析の評価と問い合わせ機構の構築			
<p>近年 Web 上での仮想組織の活動が活発になっており、Web 空間でどんな活動が行われているかを調べる ことが重要となってきた。また、個人や状況に応じて情報をパーソナライズして調べることも重要で ある。</p> <p>一方本研究室では、多次元データキューブの下でデータマイニングを行う機構であるアイテムセットキュー ブの試作を行っている。アイテムセットキューブでは各属性で指定されたセルに、条件を満たす高頻度アイ テムセットを格納している。これにより、各属性に着目した分析を即座に行うことができる。昨年著者の グループでは、このアイテムセットキューブを用い多次元制約機構の上で高頻度アイテムセットとして Web 構造解析というコア（完全 2 部グラフ）を求め、このコア単位の関係でノードを構成するコアコミュニティ グラフと呼ぶグラフ構造の作成を行った。そして、このグラフ上でノードのランキングを行うことによって 重要な組織を目立たせ、イントラネット型の Web 空間の構造解析を行った。</p> <p>この多次元制約機構を用いて、ページのリンク関係の FROM/TO の属性に着目した分析を行うことによ って、制約なしで分析を行った場合からは分からなかった結果などを得ることができ、手法の有効性を感じる ことができた。しかし、この分析で得られた結果を詳細に調べたところ、作成されたコアコミュニティグラ フにもとづきランキングした結果について、グラフのエッジの関係などの構造とランキングの間の隔たりが 大きい場合もあるという状況が見られた。</p> <p>そこで本論文では、多次元制約機構上の Web 構造解析で用いるコアコミュニティグラフとランキングに 用いる式のモデルの再検討を行った。このモデル改善により、よりグラフ構造を適切に反映したランキング 結果を得られるようにした。その上で電気通信大学ドメインのリンク構造データの分析結果から、この手法 によって、イントラネット型 Web 空間の階層構造以外の構造をどの程度捉えることができているのか、改 めて評価し報告する。</p> <p>また、より詳細に分析を行うための問い合わせ機構の構築を行った。具体的には、</p> <ul style="list-style-type: none">● 時間経過でのコミュニティの変化の調査● ホップ数の調整による制約範囲の調整● キーワードを用いた制約 <p>に対応する問い合わせを、多次元制約機構の属性として新たに取り入れ評価を行った。時間経過によるコ ミュニティ変化の調査では、Jaccard 係数を用いてノードを構成するページの変化などを調べ、時間経過に より新たに現れるコミュニティや、消えてしまうコミュニティなどといった変化を調べることができること を示した。そして、ホップ数を用いた制約範囲の調整を行うことにより、中間ページを経た参照関係まで分 析対象にすることで、従来の直接のリンク関係を用いていただけでは分からなかったコミュニティの繋がり などを分析することが可能であることを示した。またキーワードを用いた制約を用いることで、任意のペ ージ集合を対象にした分析が行えるようになった。これにより、ドメイン制約よりも適切に対象を絞り込んで 分析を行えるため、対象となる Web 空間についてより詳細に分析を行うことが可能であることを示した。</p> <p>これらの属性を組み合わせた問い合わせを行うことで、コミュニティ構造分析の場合でも詳細に条件を定 めた分析が可能であることが示せた。</p>				